TACKLING BACKGROUND DIFFERENTLY

ICVSS 2015
Sicily ~ 12-18 July
International Computer Vision Summer School

Jetley S., Romera-Paredes B., Torr P. - University of Oxford
{sjetley, bernard, phst}@robots.ox.ac.uk

Abstract

Practical vision systems need to identify objects from amongst n different classes of interest while distinguishing them from the residual image content i.e. the background. Background does not have one consistent definition, yet highly successful approaches based on CNNs[1, 2], treat it as a regular object category and attempt to learn its appearance. This is counter-intuitive.
We propose a modified deepnet architecture for tackling background class as 'something other than the objects of interest'. We filter the background samples using a threshold that is learnt via end-to-end training and not as an optimum of a sub-problem. Separating the background samples allows us to define unambiguous output embedding space for just the object classes which offers a promising potential to boost classification as well as zero-shot performance.

Motivation

Issues with the traditional way of handling 'background' in a deepnet:

- Learning **dissimilar visual patterns** as background class is counter-intuitive
- Priors such as **natural language semantics, attributes are not available** for background class

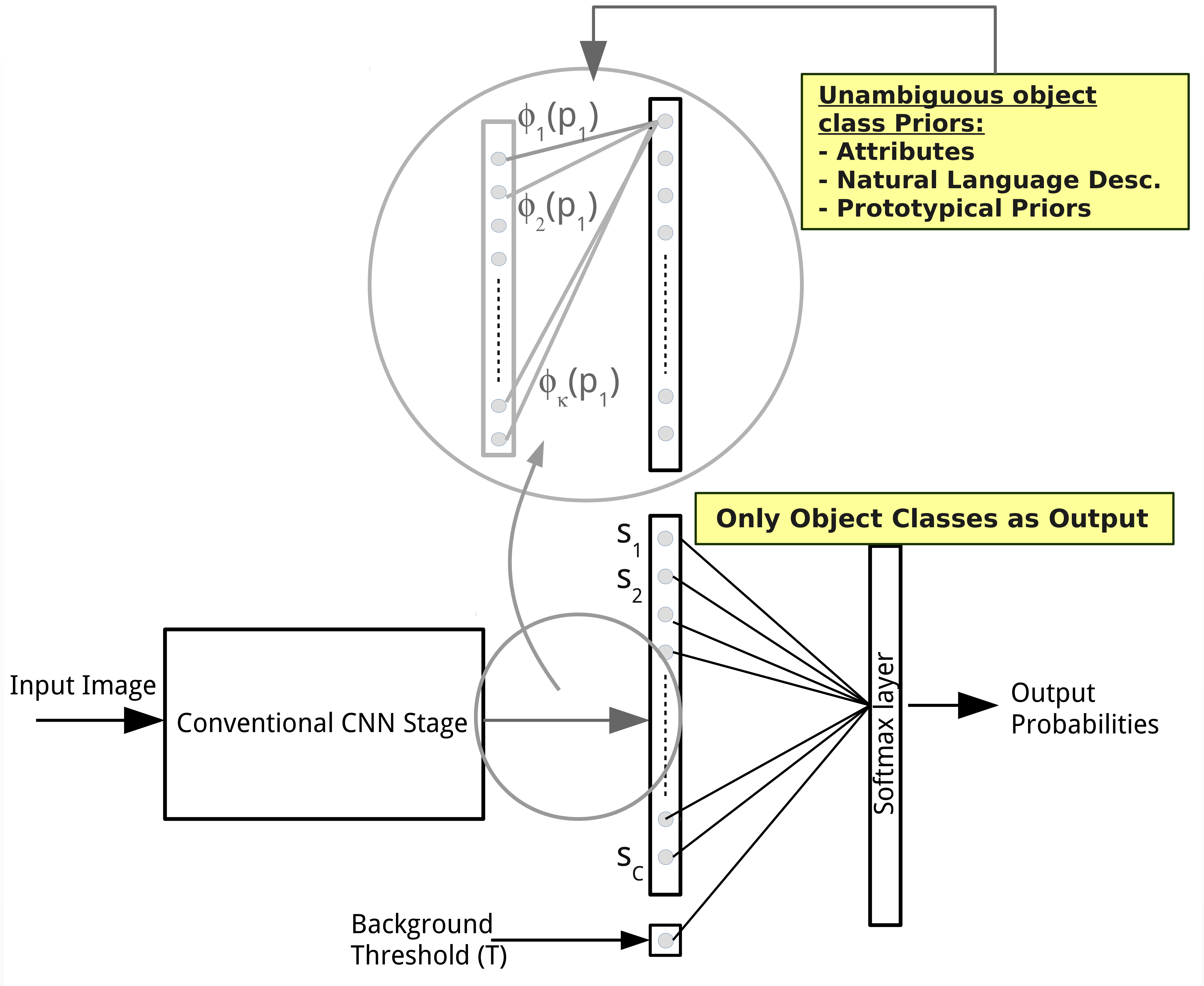
ProposedApproach

The aim is to describe the background class with a threshold T (fixed or learned), such that:

$$x = \begin{cases} \text{background}, & \text{if } s_c < T \ \forall c \in \{1,\dots,C\} \\ \text{object}, & \text{otherwise} \end{cases}$$

, where x is the input sample and s_c is the score (activation) before the final softmax layer and C is the total number of object classes (ref. Fig.1).

Proposed Architecture



Unambiguous object class Priors:
- Attributes
- Natural Language Desc.
- Prototypical Priors

Only Object Classes as Output

References

[1] Ren, Shaoqing, et al. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." arXiv:1506.01497 (2015).

[2] Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." arXiv:1411.4038 (2014).

[3] Jetley, Saumya, et al. "Prototypical Priors: From Improving Classification to Zero-Shot Learning." BMVC. 2015.

Benefits

- Background is intuitively treated as an entity other than the objects of interest, for which **object class activation is <** T
- Allows **end-to-end learning** of background threshold
- **Unambiguous class priors** such as natural language semantics or visual attributes can be leveraged in end-to-end training [3] **as the set of fixed weights** ϕ (Ref. Fig.1)

Dataset

German Traffic Sign Dataset (43 classes):
Training: 39209 object samples & 1410 background samples
Val & Test: 6316 object samples & 450 background samples; 6315 object samples & 525 background samples

Results

Results are promising:

Description	Test Acc.(%)
Background as object	95.77
Fixed-T	95.19
Learned-T	95.68